



# Full wwPDB X-ray Structure Validation Report ⓘ

Jun 28, 2016 – 07:31 PM EDT

PDB ID : 5EJS  
Title : Structure of Dictyostelium Discoideum Myosin VII MyTH4-FERM MF2 domain, mutant 1  
Authors : Planelles-Herrero, V.J.; Sirkia, H.; Sourigues, Y.; Titus, M.A.; Houdusse, A.  
Deposited on : 2015-11-02  
Resolution : 2.70 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.  
We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)  
A user guide is available at  
<http://wwpdb.org/validation/2016/XrayValidationReportHelp>  
with specific help available everywhere you see the ⓘ symbol.

---

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : unknown  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20027790  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20027790

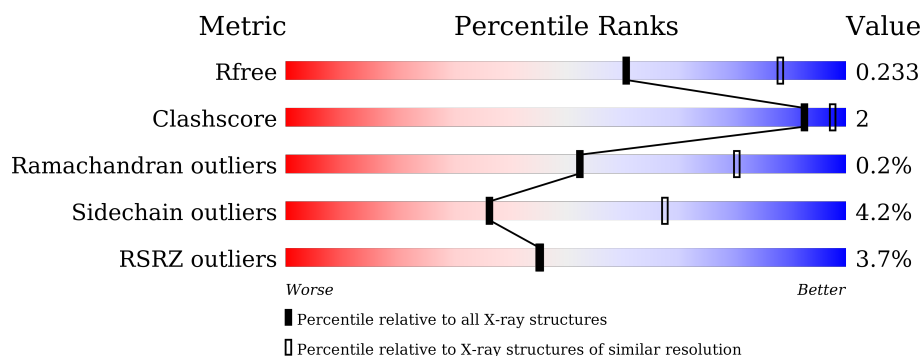
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.70 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	91344	2103 (2.70-2.70)
Clashscore	102246	2422 (2.70-2.70)
Ramachandran outliers	100387	2382 (2.70-2.70)
Sidechain outliers	100360	2382 (2.70-2.70)
RSRZ outliers	91569	2107 (2.70-2.70)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions  $\leq 5\%$ . The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	496	<div> <div>2%</div> <div> <div></div> <div>93%</div> <div>6%</div> <div>.</div> </div> </div>
1	B	496	<div> <div>5%</div> <div> <div></div> <div>89%</div> <div>7%</div> <div>..</div> </div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 8269 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Myosin-I heavy chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	496	Total	C	N	O	S	0	0	0
			4005	2554	659	770	22			
1	B	485	Total	C	N	O	S	0	0	0
			3922	2504	647	750	21			

There are 34 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	?	-	GLN	deletion	UNP Q9U1M8
A	?	-	GLN	deletion	UNP Q9U1M8
A	?	-	ALA	deletion	UNP Q9U1M8
A	?	-	THR	deletion	UNP Q9U1M8
A	?	-	LEU	deletion	UNP Q9U1M8
A	?	-	LYS	deletion	UNP Q9U1M8
A	?	-	ARG	deletion	UNP Q9U1M8
A	?	-	LYS	deletion	UNP Q9U1M8
A	?	-	ALA	deletion	UNP Q9U1M8
A	?	-	PRO	deletion	UNP Q9U1M8
A	?	-	VAL	deletion	UNP Q9U1M8
A	1909	GLU	LYS	engineered mutation	UNP Q9U1M8
A	1912	GLU	LYS	engineered mutation	UNP Q9U1M8
A	1913	GLU	LYS	engineered mutation	UNP Q9U1M8
A	2358	LEU	-	expression tag	UNP Q9U1M8
A	2359	GLU	-	expression tag	UNP Q9U1M8
A	2360	HIS	-	expression tag	UNP Q9U1M8
B	?	-	GLN	deletion	UNP Q9U1M8
B	?	-	GLN	deletion	UNP Q9U1M8
B	?	-	ALA	deletion	UNP Q9U1M8
B	?	-	THR	deletion	UNP Q9U1M8
B	?	-	LEU	deletion	UNP Q9U1M8
B	?	-	LYS	deletion	UNP Q9U1M8
B	?	-	ARG	deletion	UNP Q9U1M8
B	?	-	LYS	deletion	UNP Q9U1M8

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	?	-	ALA	deletion	UNP Q9U1M8
B	?	-	PRO	deletion	UNP Q9U1M8
B	?	-	VAL	deletion	UNP Q9U1M8
B	1909	GLU	LYS	engineered mutation	UNP Q9U1M8
B	1912	GLU	LYS	engineered mutation	UNP Q9U1M8
B	1913	GLU	LYS	engineered mutation	UNP Q9U1M8
B	2359	LEU	-	expression tag	UNP Q9U1M8
B	2360	GLU	-	expression tag	UNP Q9U1M8
B	2361	HIS	-	expression tag	UNP Q9U1M8

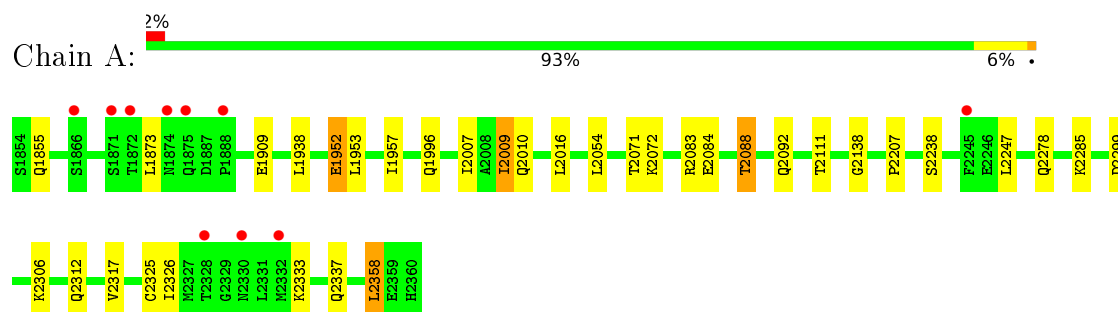
- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	202	Total O 202 202	0	0
2	B	140	Total O 140 140	0	0

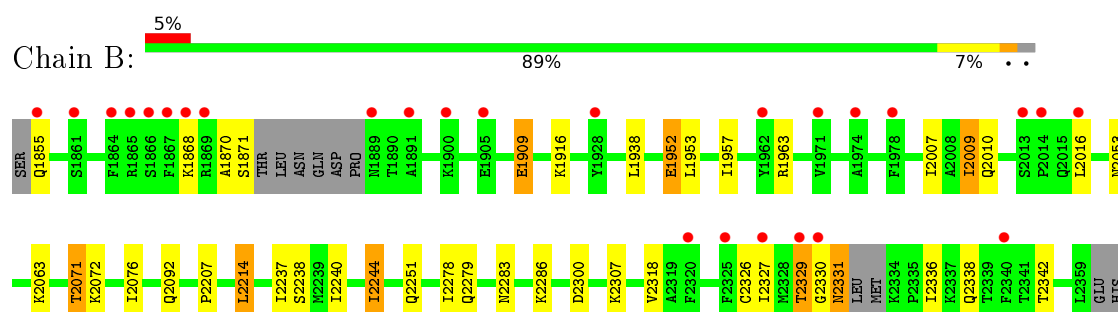
### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ( $RSRZ > 2$ ). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1: Myosin-I heavy chain



- Molecule 1: Myosin-I heavy chain



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	40.17Å 158.84Å 194.66Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	38.91 – 2.70 38.91 – 2.70	Depositor EDS
% Data completeness (in resolution range)	99.2 (38.91-2.70) 99.7 (38.91-2.70)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.09	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.21 (at 2.69Å)	Xtriage
Refinement program	BUSTER-TNT	Depositor
R, $R_{free}$	0.205 , 0.235 0.205 , 0.233	Depositor DCC
$R_{free}$ test set	1756 reflections (5.26%)	DCC
Wilson B-factor (Å <sup>2</sup> )	59.6	Xtriage
Anisotropy	0.415	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 59.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	8269	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	79.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.33% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

## 5 Model quality

### 5.1 Standard geometry

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 5$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z  >5	RMSZ	# Z  >5
1	A	0.51	0/4087	0.67	0/5526
1	B	0.50	0/4002	0.68	0/5408
All	All	0.50	0/8089	0.67	0/10934

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 5.2 Too-close contacts

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	4005	0	3967	12	0
1	B	3922	0	3892	20	0
2	A	202	0	0	0	0
2	B	140	0	0	1	0
All	All	8269	0	7859	32	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 2.

All (32) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:2331:ASN:O	1:B:2331:ASN:ND2	1.93	1.02

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:2329:THR:OG1	1:B:2336:ILE:O	2.09	0.71
1:B:2329:THR:OG1	1:B:2336:ILE:HG23	1.93	0.69
1:A:2312:GLN:HE22	1:A:2358:LEU:HD12	1.67	0.59
1:A:2317:VAL:HG12	1:A:2326:ILE:HG22	1.85	0.57
1:B:2318:VAL:HG12	1:B:2327:ILE:HG22	1.86	0.56
1:B:2063:LYS:HE2	1:B:2071:THR:HG21	1.88	0.54
1:B:1952:GLU:H	1:B:1952:GLU:CD	2.13	0.52
1:A:1952:GLU:CD	1:A:1952:GLU:H	2.14	0.51
1:A:2009:ILE:HD13	1:A:2016:LEU:HD23	1.93	0.50
1:A:2299:ASP:HB2	1:A:2306:LYS:HE3	1.93	0.50
1:B:2300:ASP:HB2	1:B:2307:LYS:HE3	1.94	0.50
1:B:2009:ILE:HD13	1:B:2016:LEU:HD23	1.93	0.49
1:B:2053:ASN:HB2	2:B:2411:HOH:O	2.12	0.49
1:B:2214:LEU:HB3	1:B:2237:ILE:HG23	1.95	0.48
1:B:2240:ILE:HG22	1:B:2244:ILE:HD12	1.94	0.48
1:B:1868:LYS:C	1:B:1870:ALA:H	2.17	0.48
1:A:2007:ILE:O	1:A:2010:GLN:HG2	2.14	0.47
1:A:2072:LYS:HD2	1:A:2092:GLN:HB3	1.99	0.45
1:B:2007:ILE:O	1:B:2010:GLN:HG2	2.16	0.44
1:A:1953:LEU:O	1:A:1957:ILE:HG13	2.18	0.44
1:A:2084:GLU:O	1:A:2088:THR:HG22	2.18	0.42
1:B:2330:GLY:HA2	1:B:2331:ASN:HA	1.62	0.42
1:A:2111:THR:OG1	1:A:2138:GLY:HA3	2.20	0.42
1:B:1909:GLU:H	1:B:1909:GLU:HG2	1.63	0.42
1:B:2072:LYS:HD2	1:B:2092:GLN:HB3	2.01	0.42
1:A:2312:GLN:HE22	1:A:2358:LEU:CD1	2.33	0.41
1:B:1953:LEU:O	1:B:1957:ILE:HG13	2.20	0.41
1:B:2076:ILE:HG21	1:B:2076:ILE:HD13	1.82	0.41
1:B:2329:THR:HB	1:B:2330:GLY:H	1.71	0.41
1:A:2054:LEU:HD23	1:A:2054:LEU:HA	1.97	0.40
1:B:2278:ILE:HG22	1:B:2342:THR:HB	2.02	0.40

There are no symmetry-related clashes.

## 5.3 Torsion angles [i](#)

### 5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.



The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	492/496 (99%)	477 (97%)	14 (3%)	1 (0%)	52	80
1	B	479/496 (97%)	461 (96%)	17 (4%)	1 (0%)	52	80
All	All	971/992 (98%)	938 (97%)	31 (3%)	2 (0%)	52	80

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	2207	PRO
1	B	2207	PRO

### 5.3.2 Protein sidechains ⓘ

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	455/458 (99%)	437 (96%)	18 (4%)	38	69
1	B	445/458 (97%)	425 (96%)	20 (4%)	34	65
All	All	900/916 (98%)	862 (96%)	38 (4%)	36	68

All (38) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1855	GLN
1	A	1873	LEU
1	A	1909	GLU
1	A	1938	LEU
1	A	1952	GLU
1	A	1996	GLN
1	A	2009	ILE
1	A	2071	THR
1	A	2083	ARG
1	A	2088	THR
1	A	2238	SER

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	2247	LEU
1	A	2278	GLN
1	A	2285	LYS
1	A	2325	CYS
1	A	2333	LYS
1	A	2337	GLN
1	A	2358	LEU
1	B	1855	GLN
1	B	1871	SER
1	B	1909	GLU
1	B	1916	LYS
1	B	1938	LEU
1	B	1952	GLU
1	B	1963	ARG
1	B	2009	ILE
1	B	2071	THR
1	B	2214	LEU
1	B	2238	SER
1	B	2244	ILE
1	B	2251	GLN
1	B	2279	GLN
1	B	2283	ASN
1	B	2286	LYS
1	B	2326	CYS
1	B	2329	THR
1	B	2331	ASN
1	B	2338	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	2208	ASN
1	A	2312	GLN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

There are no ligands in this entry.

## 5.7 Other polymers [i](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues [i](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	A	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	A	1875:GLN	C	1887:ASP	N	16.94

## 6 Fit of model and data

### 6.1 Protein, DNA and RNA chains

In the following table, the column labelled ‘#RSRZ> 2’ contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95<sup>th</sup> percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled ‘Q< 0.9’ lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å <sup>2</sup> )	Q<0.9
1	A	496/496 (100%)	-0.13	10 (2%) 68 69	36, 67, 116, 205	0
1	B	485/496 (97%)	0.14	26 (5%) 29 28	38, 82, 144, 187	0
All	All	981/992 (98%)	0.00	36 (3%) 45 45	36, 72, 137, 205	0

All (36) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	2330	GLY	5.6
1	A	1888	PRO	5.4
1	B	1868	LYS	4.8
1	A	1874	ASN	4.6
1	B	1855	GLN	4.3
1	B	1861	SER	4.3
1	B	1867	PHE	3.7
1	A	1875	GLN	3.7
1	A	1872	THR	3.5
1	B	1869	ARG	3.4
1	B	2014	PRO	3.1
1	A	2330	ASN	3.1
1	B	2320	PHE	3.0
1	B	2329	THR	3.0
1	B	1905	GLU	3.0
1	B	1864	PHE	2.8
1	B	1962	TYR	2.8
1	B	2327	ILE	2.8
1	B	1889	ASN	2.7
1	B	1866	SER	2.7
1	B	1865	ARG	2.6
1	B	1971	VAL	2.5
1	B	1891	ALA	2.5
1	A	1866	SER	2.4

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	B	1928	TYR	2.3
1	B	2013	SER	2.3
1	A	1871	SER	2.3
1	B	1900	LYS	2.3
1	A	2332	MET	2.2
1	A	2328	THR	2.2
1	A	2245	PHE	2.2
1	B	2016	LEU	2.2
1	B	1978	PHE	2.1
1	B	2325	PHE	2.1
1	B	1974	ALA	2.1
1	B	2340	PHE	2.1

## 6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

## 6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 6.4 Ligands [i](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.